Introduction

Genome-wide association studies (GWAS) have been an important tool for identifying genetic variants for many traits and diseases. Disease traits are investigated through GWAS by genotyping large populations and calculating differences in SNP (single-nucleotide polymorphism) frequencies between individuals with the trait and individuals without the trait (Uffelman et al, 2021). GWAS have become increasingly more powerful, as meta-analysis methods and cheaper genotyping technology have allowed newer GWAS to include over 1 million participants (Loos, 2019).

To functionally follow up a GWAS result, researchers have investigated which cell types are important for a trait or disease. To link GWAS to cell types, researchers usually identify gene sets or gene properties that define certain cell types based on single cell RNA (scRNA) sequencing data; these datasets quantify the gene expression in each cell. This information is subsequently used to cluster cells into cell types, which are labeled based on prior biological knowledge. It can be tested whether the GWAS signal is enriched for the genes belonging specifically to one cell type. Therefore, the outcome of the above analysis depends on the quality of the GWAS and the chosen scRNA dataset.

In a previous study, an atlas of 4,155 GWASs was cataloged (Watanabe et al, 2019); this resource allows researchers to view risk SNPs identified by each GWAS and but does not yet inform on which cell types are associated with these SNPs. FUMA also provides a tool for mapping cell type to GWAS signals, but comparing multiple GWAS and their associated cell types remains difficult. Thus, our project aims to produce a tool that visualizes FUMA's functional analysis, comparing FUMA results for multiple GWAS of the same disease trait to identify genes or cell types that are implicated repeatedly. Cellular gene expression is measured in scRNA sequencing studies, which FUMA maps to the genetic expression profile observed for a disease trait investigated in a GWAS. Single-cell RNA sequencing datasets will be selected for this analysis based on their level of annotation: many popular scRNA sequencing databases for brain cell types, such as the Allen Brain human and mouse datasets, have been updated multiple times to have more specific labeling of cell types, drastically increasing the number of cell types present in each dataset. Better cell type annotation lends itself to more specific cell type identities linked with the GWAS trait or disease.

To illustrate our approach for comparing cell types associated with a disease trait we have chosen to focus on Parkinson's disease. While the phenotype for this disease is clear, there is much more to explore about which genetic variants drive Parkinson's disease and the relevant cell types that could be targeted for treatment. Parkinson's disease (PD) is characterized phenotypically by a progressive loss of autonomous motor function, which is linked to loss of dopaminergic (DA) neurons in DA-rich regions of the brain such as the midbrain (Kouli, *Parkinson's Disease*). In the brain-first progression of PD, a-synuclein aggregates, known as Lewy bodies, appear first in the medulla and olfactory bulb and contribute to toxicity-driven degeneration of DA neurons (Yamasaki et al, 2022). Parkinson's is a progressive disease, with a-synuclein driven degeneration spreading to different regions of the brain slowly over the

course of decades. PD progression is marked by Braak staging: 5 stages of worsening degeneration as the disease spreads from the substantia nigra to deeper regions of the midbrain (Dickson et al, 2010). Recent analysis of PD patient-derived induced pluripotent stem cells (iPSCs) has confirmed the accuracy of Braak staging and the implicated brain regions (Fernandes et al, 2020).

Treatment for Parkinson's Disease has been primarily focused on restoring healthy dopamine signaling (Armstrong et al, 2020), therefore further study into specific neuron and glia cell types implicated in PD pathology is warranted for improved targeted treatment on dopamine populations. We will address our research question: which genomic risk loci and cell types repeatedly emerge from PD GWAS? We can hopefully assert that the genomic loci and cell types that map most consistently to PD GWAS are the most associated with PD.

In this study, we first tested if cell types identified from multiple PD GWAS and different scRNAseq datasets converge. Next, we developed a tool to explore genomic risk loci and cell types that are associated with PD, with interactive elements that allow researchers to navigate cell types prioritized by different PD GWAS and scRNAseq datasets. This tool is aimed to help future researchers catalog which cell types are relevant to their disease of choice (and potentially to other traits of interest), and the pipeline for visualizing this overlap will be demonstrated using GWAS for Parkinson's disease.

Methods

To access GWAS summary statistics for Parkinson's disease (PD), the GWAS Atlas platform was utilized (https://atlas.ctglab.nl/). Nine studies are available on the GWAS Atlas for non-familial PD; we selected the four latest GWAS to analyze, excluding the five older studies for having too small of a population genotyped which results in no genome-wide significant SNPs, or for having incomplete data. The four GWAS selected (Supplementary Table 1) were Nalls et al (2019), Blauwendraat et al (2019), Pankratz et al (2012), and Simon et al (2009).

Pre-processing of GWAS summary statistics entailed two steps. First, SNP uniqIDs were split into separate columns of chromosome and position for each SNP. FUMA's SNP2GENE tool requires these columns to be separate in order to recognize chromosome and position data. Next, pre-processed summary statistics were validated by checking the length of the file, and ensuring that it matched the line count for the same GWAS on the GWAS Atlas.

Using the SNP2GENE results, we first assembled a comparison of the overlap between genomic risk loci identified for Parkinson's disease. SNP2GENE identifies genomic risk loci using multiple "LD blocks", regions with largely self-contained linkage disequilibrium which result in associated gene expression. LD blocks vary widely in size, but within the LD block there is expected to be multiple candidate SNPs that are associated with a select few "lead SNPs" which potentially drive the phenotype (Watanabe et al, 2019). Since LD blocks can be of any size, we defined overlap in genomic risk loci as any genomic risk loci that overlap at any

location within start to end. This comparison of genomic risk loci was plotted using an Upset diagram and overlapping loci were identified.

We next compared the nearest genes associated with significant SNPs. These genes are selected simply as those in the closest proximity to the SNPs that fall within the genome-wide significance threshold of 5*10^-8 (Watanabe et al, 2019). We compared overlap between GWAS to identify nearest genes that repeatedly emerge in association with PD. Overlap in nearest genes was plotted using an Upset diagram and overlapping genes were listed.

Lastly, we compared cell type associations with Parkinson's disease. Gene expression at a single-cell resolution is used to find cell types which express genes associated with GWAS traits. Mapping of GWAS to cell types is performed using MAGMA gene-property analysis, which relies on a multiple regression approach (Watanabe et al, 2019); this is implemented in FUMA's Cell Type tool. We used 31 scRNAseq datasets, which each capture different regions of the brain. We compared overlap in cell types associated with PD and plotted them using an Upset diagram. Additionally, we compared cell types identified in this analysis with PD cell types identified in previous studies and we also further investigated whether any cell types from unique brain regions (based on the scRNAseq dataset) were identified.

To simplify these comparison tests, we created an application using RShiny to compare overlap in genomic risk loci, nearest genes, and cell types between multiple GWAS for the same trait. The goal of this app is to supplement FUMA results with a comparison tool, so that the user can download their FUMA results and use our pipeline to aggregate their results for multiple GWAS and easily visualize the comparison between GWAS.

The key feature of this tool is the identification of cell types that repeatedly emerge in association with PD. To this end, the first function of our Shiny app is a barplot that lists cell types and their p-values, denoting the significance of their association with PD. Using this barplot, the user can identify the names of cell types that are considered significant for PD. The user could also observe certain groups of cell types that repeatedly emerge, such as glial cells or excitatory neurons.

Finally, our app also describes the pipeline the user will need to follow in order to compare multiple GWAS. Users must first download their GWAS summary statistics from the GWAS Atlas, then use our pre-processing script to prepare their summary statistics, then run their summary statistics on SNP2GENE and Cell type on the FUMA platform. They can then use our Shiny script to aggregate the FUMA results and visualize them in the Shiny app.

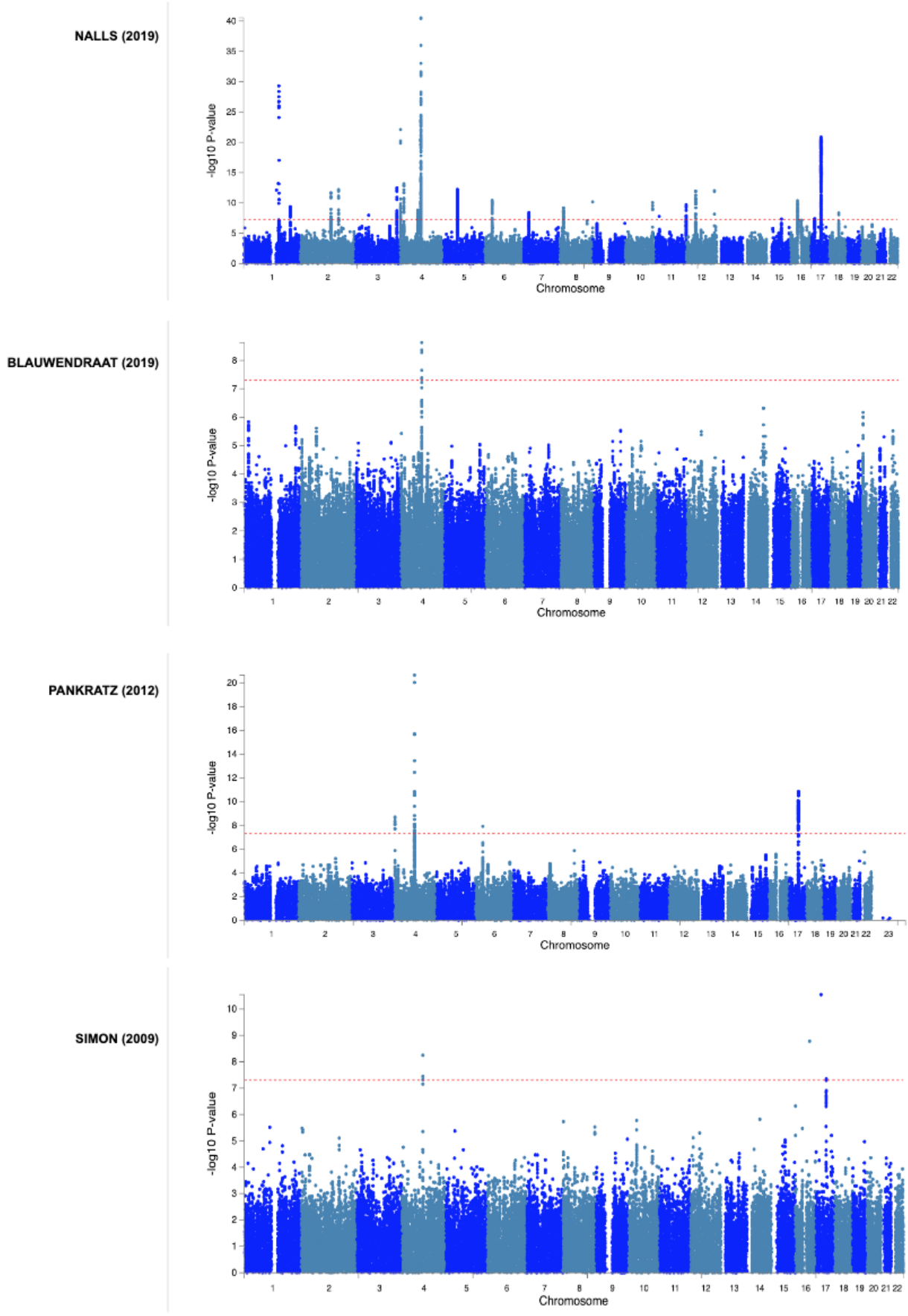


Figure 1: Manhattan Plots of the four Parkinson's disease GWAS selected for this study. SNPs are plotted as dots in the Manhattan plot. Chromosomal position is shown on the x-axis, and log-adjusted P-values are shown on the y-axis. The genome-wide significance threshold of 5*10^-8 is displayed as a dotted red line; all significant SNPs fall above this significance threshold.

Results

Across PD GWAS, there is a concordant identification of significant SNPs within chromosome 4, and some repeated identification of significant SNPs in chromosome 17 (Fig. 1). Nalls et al (2019) uniquely identifies multiple significant SNPs in chromosome 1 risk loci. However, although Nalls et al (2019) utilizes the highest number of samples, it also captures a very low heritability when compared to other PD GWAS (Fig. 2).

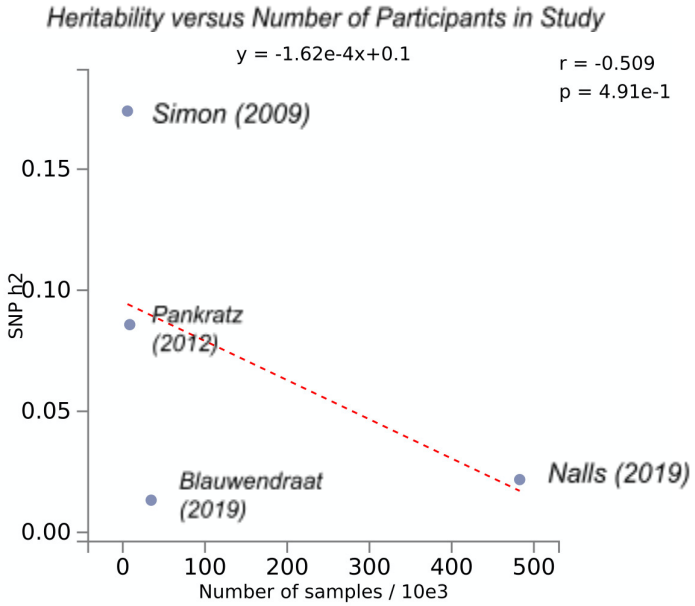
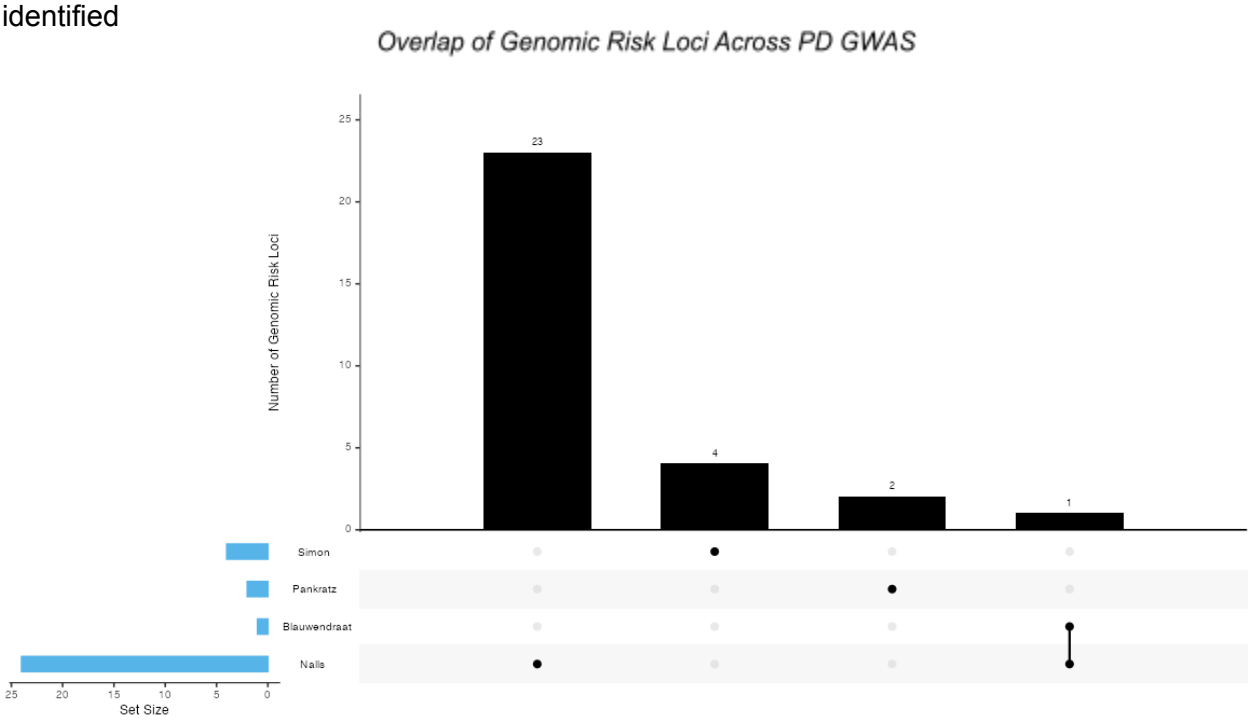


Figure 2: Scatterplots of the four Parkinson's disease GWAS selected for this study describing the relationship between number of participants in the study (showing the size of the study) versus the heritability (h^2) of Parkinson's disease captured by this study. SNP heritability is in liability scale (Watanabe et al, 2019). Number of participants is shown on the x-axis, and heritability is shown on the y-axis. Each GWAS's name is listed on the plot.

We also tested if genomic risk loci overlapped and found only one genomic risk locus in chromosome 4 which was identified by more than one GWAS (Fig. 3). A total of 30 genomic risk loci were identified between all four Parkinson's disease GWAS, located primarily in chromosomes 1, 4, and 17 (Supplementary Table 2). Of these 30 risk loci, 29 were unique to GWAS while one risk locus containing the SNP with rsID 356203 (located within chromosome 4) was identified by both Blauwendraat et al (2019) and Nalls et al (2019). Overall, there is little overlap of genomic risk loci between Parkinson's disease GWAS, and investigating the locus that did repeatedly emerge did not result in a novel identification of SNPs associated with PD.

There is a much stronger overlap between genes nearest to SNPs significant for PD, with approximately forty genes in chromosome 4 and 17 being implicated by multiple GWAS (Fig. 4). SNCA, a gene that codes for the protein alpha-synuclein, is implicated by all four GWAS. Other genes that emerge from at least 2 GWAS include genes that code for retinal ganglion cells, tau

phosphorylation genes, and corticotropin-releasing hormone genes (Fig. 4). A total of 110 genes were identified as those nearest to significant SNPs; of these 110 genes, 100 were uniquely identified

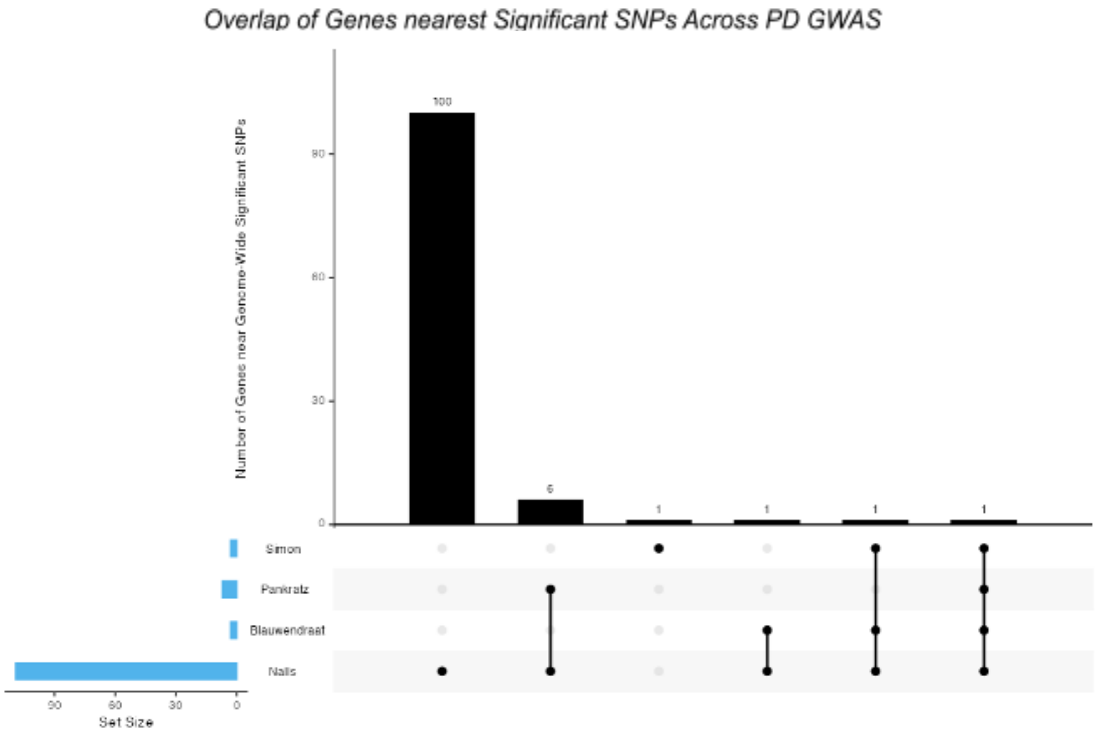


rsID	Nalls	Pankratz	Simon	Blauwendraat	sum
rs356203	1	0	0	1	2

Figure 3: Upset diagram of the overlap in genomic risk loci identified by the four Parkinson's disease GWAS. Number of genomic risk loci is shown on the y-axis, and GWAS is shown on the y-axis. Genomic risk loci identified by multiple GWAS appear not as singular dots, but as a line connecting multiple dots.

by Nalls et al (2019), 1 was uniquely identified by Simon et al (2009), and 9 genes were identified from multiple GWAS. Therefore, overlap between PD GWAS at the gene level seems to be more concordant than overlap between individual SNPs.

Next, we investigated overlap between cell types associated with PD. We first identified all cell types that were considered significant after multiple testing correction with Bonferroni. The number of cell types within each scRNAseq dataset was used as the number of tests within multiple testing correction; the resulting value was called "P.adj.pds", which stands for P-adjusted within the scRNAseq dataset. We also applied a log10 transformation and ranked the cell types to identify those that were most associated with PD GWAS signal (Fig. 5). We identified a total of 20 significant cell types (Table 1), but each of these cell types were identified by a single GWAS. There is no overlap in significant cell types from multiple GWAS (Fig. 6). However, there is repeated implication of the same cell type groups. From different GWAS, excitatory neurons emerge repeatedly (vGLUT neurons, NF cortical neurons, Gad2 neurons). Oligodendrocyte, OPCs (oligodendrocyte precursors), and radial glia also appear as significant cell types but are only uniquely identified by Blauwendraat.



nearestGene	Nalls	Pankratz	Blauwendraat	Simon	sum
SNCA	1.00	1.00	1.00	1.00	4.00
RP11-115D19.1	1.00	0.00	1.00	1.00	3.00
CRHR1	1.00	1.00	0.00	0.00	2.00
CRHR1-IT1:CRHR1:RP11-105N13.4	1.00	1.00	0.00	0.00	2.00
CRHR1:RP11-105N13.4	1.00	1.00	0.00	0.00	2.00
KANSL1	1.00	1.00	0.00	0.00	2.00
MAPT	1.00	1.00	0.00	0.00	2.00
MAPT-AS1	1.00	1.00	0.00	0.00	2.00
RP11-67M1.1	1.00	0.00	1.00	0.00	2.00

Figure 4: Upset diagram of the overlap in genes nearest significant SNPs identified by the four Parkinson's disease GWAS. Number of genes is shown on the y-axis, and GWAS is shown on the x-axis. Genes identified by multiple GWAS appear not as singular dots, but as a line connecting multiple dots.

To visualize our results, the Shiny tool used in this project creates multiple Upset diagrams describing overlap, a barplot to list significant cell types, and a table containing all significant cell types. The user interface for this Shiny app is shown below (Fig. 7).

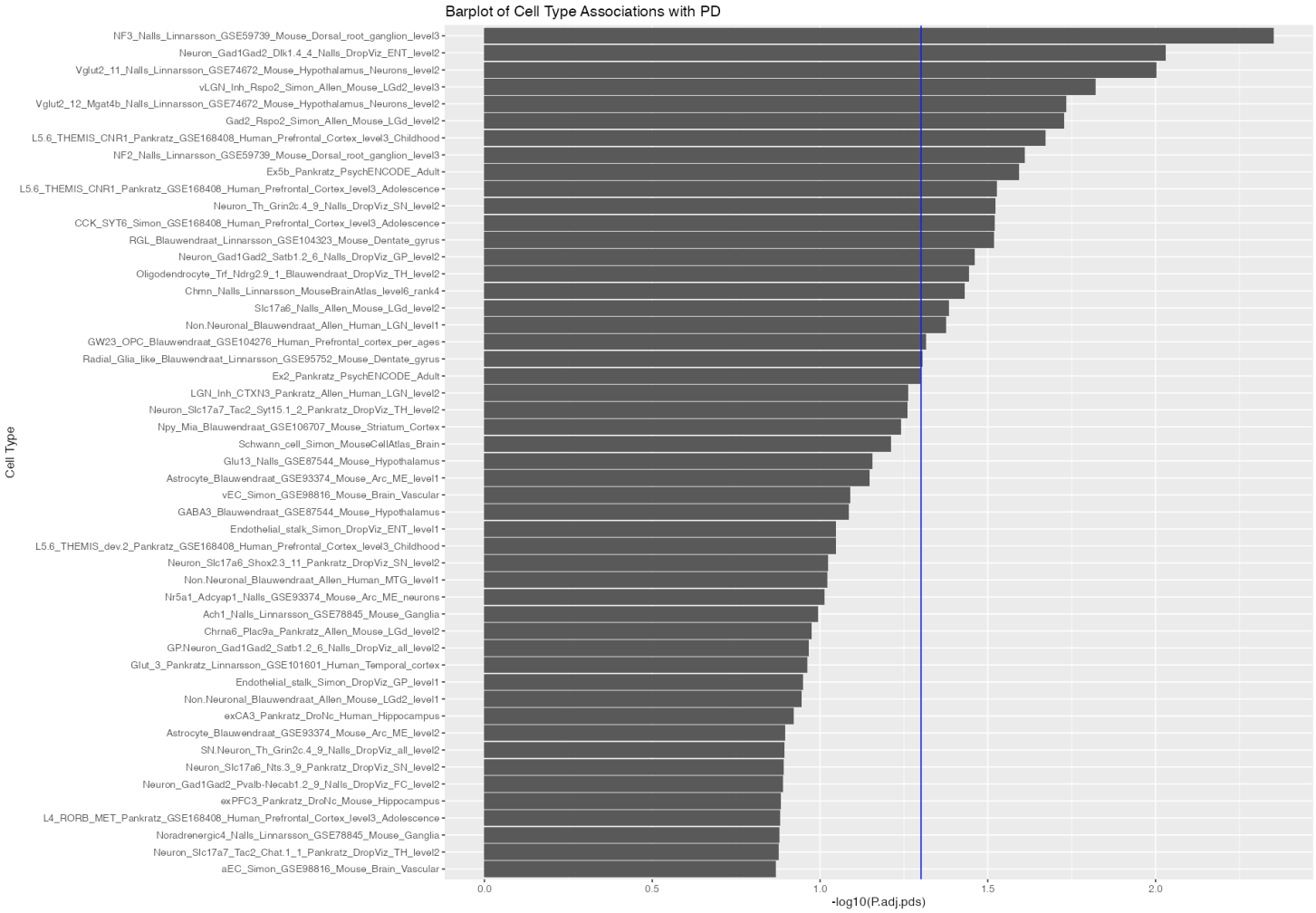


Figure 5: Barplot of cell types identified by the four Parkinson's disease GWAS. Cell type name and the GWAS and dataset it originates from is shown on the y-axis, and log-adjusted p-values are shown on the x-axis. The genome-wide significance threshold of 5*10^-8 is displayed as a blue dotted line. All significant cell types fall to the right of this significance threshold.

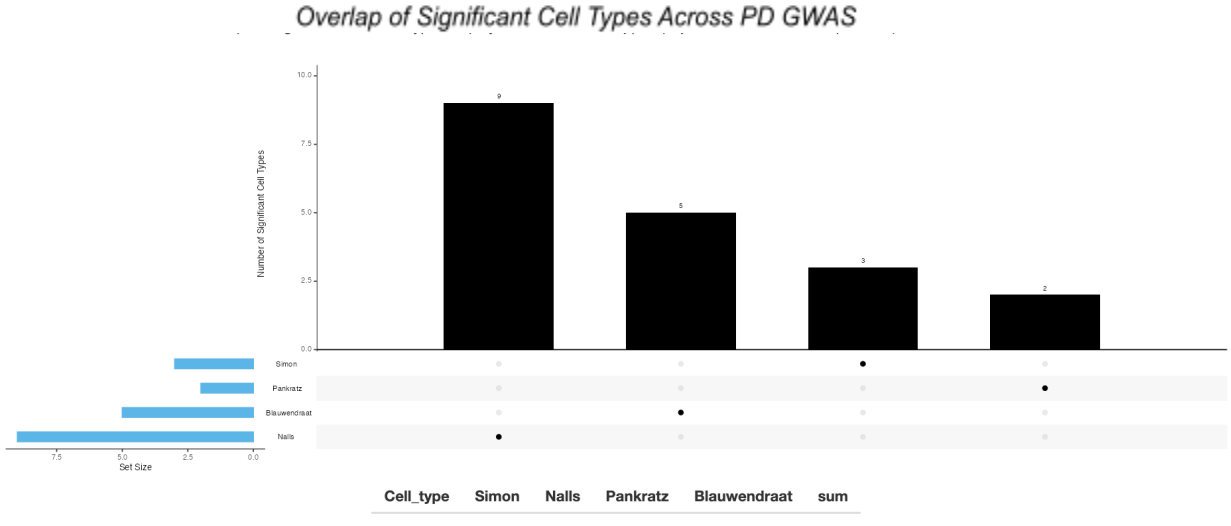


Figure 6: Upset diagram of the overlap in significant cell types identified by the four Parkinson's disease GWAS. Number of genes is shown on the y-axis, and GWAS is shown on the x-axis. Cell types identified by multiple GWAS appear not as singular dots, but as a line connecting multiple dots. There were no repeated cell types identified.

Dataset	Cell_type	logP.adj.pds	GWAS	BETA_STD	SE	P.adj.pds
Linnarsson_GSE59739_Mouse_Dorsal_root_ganglion_level3	NF3	2.350178	Nalls	0.079796	0.0094077	0.00446501
DropViz_ENT_level2	Neuron_Gad1Gad2_Dlk1.4_4	2.028348	Nalls	0.055650	0.0058997	0.00936810
Linnarsson_GSE74672_Mouse_Hypothalamus_Neurons_level2	Vglut2_11	2.002269	Nalls	0.066240	0.0197790	0.00994790
Allen_Mouse_LGd2_level3	vLGN_Inh_Rspo2	1.819956	Simon	0.086667	0.0103590	0.01513714
Linnarsson_GSE74672_Mouse_Hypothalamus_Neurons_level2	Vglut2_12_Mgat4b	1.733684	Nalls	0.046355	0.0212670	0.01846360
Allen_Mouse_LGd_level2	Gad2_Rspo2	1.727001	Simon	0.076191	0.0114280	0.01874990
GSE168408_Human_Prefrontal_Cortex_level3_Childhood	L5.6_THEMIS_CNR1	1.671238	Pankratz	0.078930	0.0367500	0.02131875
Linnarsson_GSE59739_Mouse_Dorsal_root_ganglion_level3	NF2	1.609446	Nalls	0.089371	0.0134910	0.02457840
PsychENCODE_Adult	Ex5b	1.591845	Pankratz	0.097402	0.0952280	0.02559500
GSE168408_Human_Prefrontal_Cortex_level3_Adolescence	L5.6_THEMIS_CNR1	1.526423	Pankratz	0.078228	0.0237320	0.02975616
DropViz_SN_level2	Neuron_Th_Grin2c.4_9	1.522063	Nalls	0.061436	0.0068175	0.03005637
GSE168408_Human_Prefrontal_Cortex_level3_Adolescence	CCK_SYT6	1.519529	Simon	0.069163	0.0345340	0.03023232
Linnarsson_GSE104323_Mouse_Dentate_gyrus	RGL	1.517561	Blauwendraat	0.041087	0.0978210	0.03036960
DropViz_GP_level2	Neuron_Gad1Gad2_Satb1.2_6	1.459557	Nalls	0.065556	0.0072735	0.03470907
DropViz_TH_level2	Oligodendrocyte_Trf_Ndrg2.9_1	1.443269	Blauwendraat	0.039579	0.0048816	0.03603552
Linnarsson_MouseBrainAtlas_level6_rank4	Chmn	1.430773	Nalls	0.048542	0.0287470	0.03708744
Allen_Mouse_LGd_level2	Slc17a6	1.383957	Nalls	0.056702	0.0091758	0.04130880
Allen_Human_LGN_level1	Non.Neuronal	1.374955	Blauwendraat	0.030580	0.0091413	0.04217400
GSE104276_Human_Prefrontal_cortex_per_ages	GW23_OPC	1.315600	Blauwendraat	0.055607	0.0102820	0.04835040
Linnarsson_GSE95752_Mouse_Dentate_gyrus	Radial_Glia_like	1.303678	Blauwendraat	0.032338	0.0616330	0.04969610

Table 1: This table lists all the cell types that are significant for Parkinson's disease, identified from our four Parkinson's disease GWAS. P.adj.pds refers to p-value corrected with Bonferroni within scRNAseq dataset, logP.adj.pds refers to the log-adjusted version of P.adj.pds, BETA_STD refers to the standardized beta, and SE refers to the standard error.

Discussion

In this project, we developed a tool to investigate genomic loci and cell types associated with PD. Our goal was to identify risk loci and cell types that repeatedly emerged from functional follow-up analysis of PD GWAS. However, our results suggest that PD GWAS demonstrate little concordance of genomic risk loci or of cell types, largely due to a lack of robust PD GWAS to compare. However, there is a moderate overlap of PD-associated genes. Genes nearest significant SNPs associated strongly with cell type functions (such as corticotropin release or tau phosphorylation) that have been shown to play a large role in PD (Zhang et al, 2018; Rissman 2007).

Genomic risk loci identified for PD GWAS did not show overlap across GWAS, but the risk loci identified did point to some interesting results. Most of the risk loci identified were in chromosome 4 and chromosome 17, and were mostly linked with the alpha-synuclein gene SNCA in chromosome 4 or the tau phosphorylation gene MAPT in chromosome 17 (Supplementary Table 2). Nalls et al (2019) identified 1 unique risk loci in chromosome 1 (Supplementary Table 2); this loci could be linked with the PARK family of genes. PARK genes

have been studied for their link with monogenic parkinsonism; PARK mutations are linked with early-onset parkinsonism symptoms such as motor and speech deficits (Bonifati et al, 2013). However, chromosome 1 risk loci were only identified in Nalls et al (2019) and therefore lack the validity of association with PD that chromosome 4 and chromosome 17 risk loci have.

Across PD GWAS there was limited overlap in identified cell types. While PD maps poorly to single-cell expression, certain brain regions seem to be repeatedly associated with PD. This is shown in the multiple significant cell types found in the hypothalamus and dorsal root ganglia, in both mouse and human scRNAseq datasets (Table 1). These brain regions are strongly implicated in Parkinson's disease pathology (Kouli, 2018). Furthermore, the genes we identified (Fig. 4) as associated with PD (such as RP11 genes which code for retinal ganglion cells in the optic bulb) are also validated by PD pathology; the brain-first pathology of PD is also believed to originate in the optic bulb. From our results we can conclude there is more overlap between PD GWAS at the brain region level rather than at the cell type level.

However, identifying overlap in cell types is further complicated by the convention of cell type naming in scRNAseq datasets and the lack of standard naming across different datasets. Usage of multiple scRNAseq datasets in comparison is highly impacted by a lack of naming standards (Crow et al, 2018). In this study, multiple concordant cell types had to be renamed because of spelling differences and acronym usage. Furthermore, cell type annotations are highly specific, differentiating cell types with specific protein expression (i.e. MAPT vs MAPT-AS1 cells). Therefore, naming of cell types seems to play a very large role in limiting our identification of concordant cell types. In this study, we did attempt to standardize cell names within our datasets, but it remains a complicated process.

This study does identify some cell type associations with PD that could be further investigated. First, glial cells are weakly associated to PD when mapped with GWAS (Table 1), unlike with in vivo studies of PD (Smajic et al, 2022), suggesting that glial cells populations may play less of a causal role in PD and instead are reactive to the initial disease. It has been shown though that glia cell death exacerbates the disease (Olsen et al, 2021). Astrocytes and microglia have been previously linked to PD, but were largely absent from our list of significant cell types (Copas et al, 2021). Instead, our results show that NF cortical cells and OPCs were implicated repeatedly as cell types significant for PD, and both cell types play a role in the myelination of nerves (Dean et al, 2016). Yet PD pathology shows little effect on myelination; progressive demyelination does not present itself in PD as it does in other neurodegenerative diseases (Le et al, 2023).

As a tauopathy, PD may also have a more complicated cell pathology because of multiple types of cell-to-cell transmission of tau (Gibbons et al, 2019). Tauopathies result in highly heterogeneous pathways of disease progression at the cellular level (Chung et al, 2021), so while brain regions affected by PD are very clearly marked in the disease progression, the initial emergence of tau or sudden acceleration of tau tangles could occur in multiple possible cell types within those brain regions.

We conclude that there is moderate overlap in genes associated with PD across GWAS, but very minimal overlap in genomic risk loci and cell types across GWAS. Our limitations for this study include a lack of large PD GWAS, poor naming standards of scRNAseq cell types, and the complicated nature of Parkinson's and other tauopathies. However, the strength of this study is that the pipeline for comparison of multiple GWAS for the same trait could be extended to other neurological disorders similar to Parkinson's, such as insomnia, in order to further investigate other cell types that could contribute to PD.

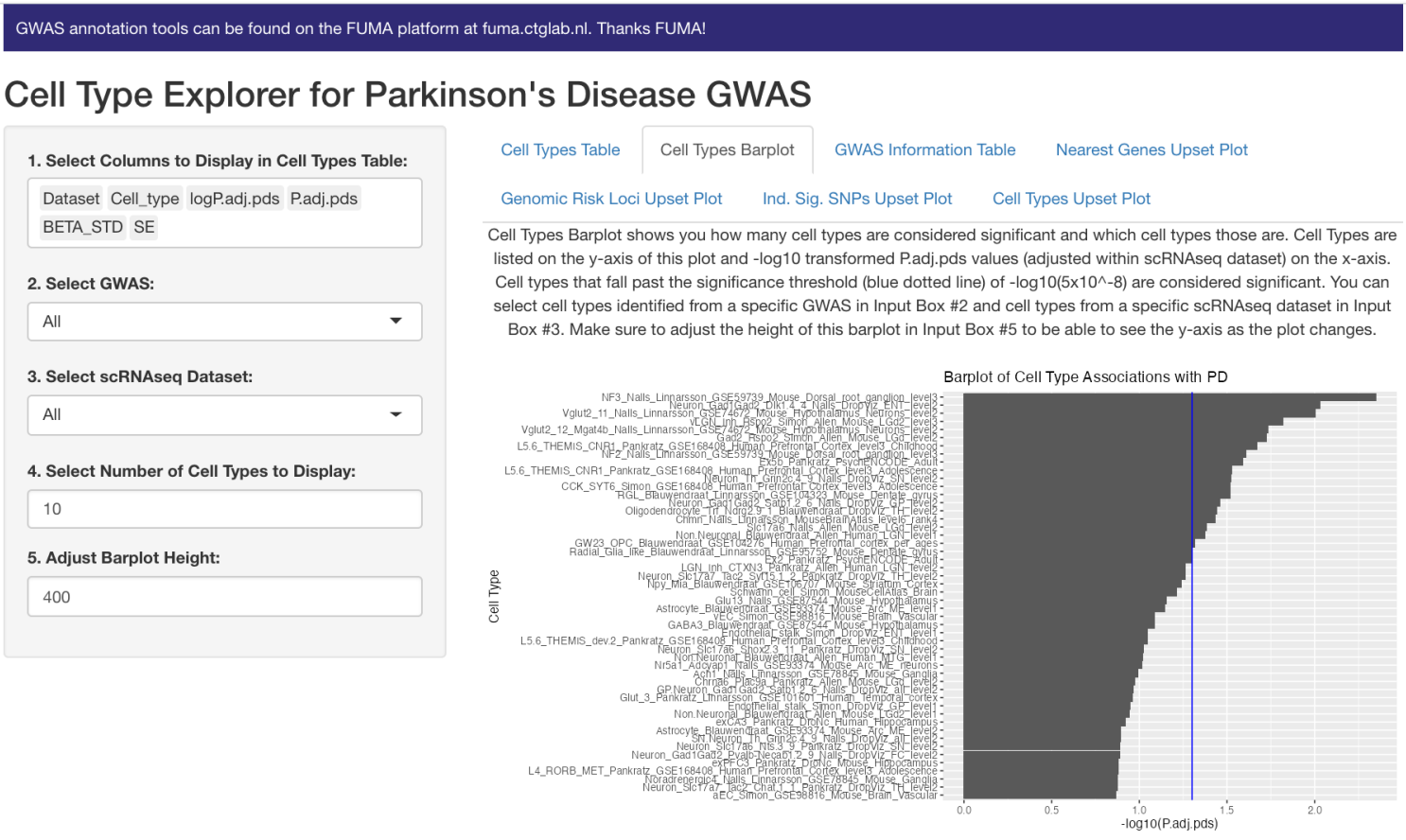


Fig 7: This is the user interface display for the GWAS comparison tool developed in this project using the Shiny package for R. This GWAS comparison tool displays overlap in genomic risk loci, nearest genes, independent significant SNPs, and cell types, across multiple PD GWAS.

References

1. Armstrong MJ, Okun MS. Diagnosis and Treatment of Parkinson Disease: A Review. JAMA. 2020 Feb 11;323(6):548-560. doi: 10.1001/jama.2019.22360. PMID: 32044947.
2. Blauwendraat C, Heilbron K, Vallerga CL, Bandres-Ciga S, von Coelln R, Pihlstrøm L, Simón-Sánchez J, Schulte C, Sharma M, Krohn L, Siitonen A, Iwaki H, Leonard H, Noyce AJ, Tan M, Gibbs JR, Hernandez DG, Scholz SW, Jankovic J, Shulman LM, Lesage S, Corvol JC, Brice A, van Hilten JJ, Marinus J; 23andMe Research Team; Eerola-Rautio J, Tienari P, Majamaa K, Toft M, Grosset DG, Gasser T, Heutink P, Shulman JM, Wood N, Hardy J, Morris HR, Hinds DA, Gratten J, Visscher PM, Gan-Or Z, Nalls MA, Singleton AB; International Parkinson's Disease Genomics Consortium (IPDGC). Parkinson's disease age at onset genome-wide association study: Defining heritability, genetic loci, and α-synuclein mechanisms. Mov Disord. 2019 Jun;34(6):866-875. doi: 10.1002/mds.27659. Epub 2019 Apr 7. PMID: 30957308; PMCID: PMC6579628.
3. Bonifati V. Genetics of Parkinson's disease--state of the art, 2013. Parkinsonism Relat Disord. 2014 Jan;20 Suppl 1:S23-8. doi: 10.1016/S1353-8020(13)70009-9. PMID: 24262182.
4. Chong CW, Lim SY, Yap IKS, Teh CSJ, Loke MF, Song SL, Tan JY, Ang BH, Tan YQ, Kho MT, Bowman J, Mahadeva S, Yong HS, Lang AE. Gut Microbial Ecosystem in Parkinson Disease: New Clinicobiological Insights from Multi-Omics. Ann Neurol. 2021 Mar;89(3):546-559. doi: 10.1002/ana.25982. Epub 2021 Jan 11. PMID: 33274480.
5. Dickson DW. Parkinson's disease and parkinsonism: neuropathology. Cold Spring Harb Perspect Med. 2012 Aug 1;2(8):a009258. doi: 10.1101/cshperspect.a009258. PMID: 22908195; PMCID: PMC3405828.
6. Gibbons CH. Basics of autonomic nervous system function. Handb Clin Neurol. 2019;160:407-418. doi: 10.1016/B978-0-444-64032-1.00027-8. PMID: 31277865.
7. Kouli A, Torsney KM, Kuan WL. Parkinson’s Disease: Etiology, Neuropathology, and Pathogenesis. In: Stoker TB, Greenland JC, editors. Parkinson’s Disease: Pathogenesis and Clinical Aspects [Internet]. Brisbane (AU): Codon Publications; 2018 Dec 21. Chapter 1. PMID: 30702842.
8. Munz M, Richter GM, Loos BG, Jepsen S, Divaris K, Offenbacher S, Teumer A, Holtfreter B, Kocher T, Bruckmann C, Jockel-Schneider Y, Graetz C, Ahmad I, Staufenbiel I, van der Velde N, Uitterlinden AG, de Groot LCPGM, Wellmann J, Berger K, Krone B, Hoffmann P, Laudes M, Lieb W, Franke A, Erdmann J, Dommisch H, Schaefer AS. Meta-analysis of genome-wide association studies of aggressive and chronic periodontitis identifies two novel risk loci. Eur J Hum Genet. 2019 Jan;27(1):102-113.
   doi: 10.1038/s41431-018-0265-5
9. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, Tan M, Kia DA, Noyce AJ, Xue A, Bras J, Young E, von Coelln R, Simón-Sánchez J, Schulte C, Sharma M, Krohn L, Pihlstrøm L, Siitonen A, Iwaki H, Leonard H, Faghri F, Gibbs JR, Hernandez DG, Scholz SW, Botia JA, Martinez M, Corvol JC, Lesage S, Jankovic J, Shulman LM, Sutherland M, Tienari P, Majamaa K, Toft M, Andreassen OA, Bangale T,

Brice A, Yang J, Gan-Or Z, Gasser T, Heutink P, Shulman JM, Wood NW, Hinds DA, Hardy JA, Morris HR, Gratten J, Visscher PM, Graham RR, Singleton AB; 23andMe Research Team; System Genomics of Parkinson's Disease Consortium; International Parkinson's Disease Genomics Consortium. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet Neurol. 2019 Dec;18(12):1091-1102. doi: 10.1016/S1474-4422(19)30320-5. PMID: 31701892; PMCID: PMC8422160.

10. Pankratz N, Beecham GW, DeStefano AL, Dawson TM, Doheny KF, Factor SA, Hamza TH, Hung AY, Hyman BT, Ivinson AJ, Krainc D, Latourelle JC, Clark LN, Marder K, Martin ER, Mayeux R, Ross OA, Scherzer CR, Simon DK, Tanner C, Vance JM, Wszolek ZK, Zabetian CP, Myers RH, Payami H, Scott WK, Foroud T; PD GWAS Consortium. Meta-analysis of Parkinson's disease: identification of a novel locus, RIT2. Ann Neurol. 2012 Mar;71(3):370-84. doi: 10.1002/ana.22687. PMID: 22451204; PMCID: PMC3354734.
11. Simón-Sánchez J, Schulte C, Bras JM, Sharma M, Gibbs JR, Berg D, Paisan-Ruiz C, Lichtner P, Scholz SW, Hernandez DG, Krüger R, Federoff M, Klein C, Goate A, Perlmutter J, Bonin M, Nalls MA, Illig T, Gieger C, Houlden H, Steffens M, Okun MS, Racette BA, Cookson MR, Foote KD, Fernandez HH, Traynor BJ, Schreiber S, Arepalli S, Zonozi R, Gwinn K, van der Brug M, Lopez G, Chanock SJ, Schatzkin A, Park Y, Hollenbeck A, Gao J, Huang X, Wood NW, Lorenz D, Deuschl G, Chen H, Riess O, Hardy JA, Singleton AB, Gasser T. Genome-wide association study reveals genetic risk underlying Parkinson's disease. Nat Genet. 2009 Dec;41(12):1308-12. doi: 10.1038/ng.487. Epub 2009 Nov 15. PMID: 19915575; PMCID: PMC2787725.
12. Uffelmann, E., Huang, Q.Q., Munung, N.S. *et al.* Genome-wide association studies. *Nat Rev Methods Primers* 1, 59 (2021). https://doi.org/10.1038/s43586-021-00056-9
13. Watanabe, K., Stringer, S., Frei, O. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet* 51, 1339–1348 (2019). https://doi.org/10.1038/s41588-019-0481-0
14. Yamasaki TR, Holmes BB, Furman JL, Dhavale DD, Su BW, Song ES, Cairns NJ, Kotzbauer PT, Diamond MI. Parkinson's disease and multiple system atrophy have distinct α-synuclein seed characteristics. J Biol Chem. 2019 Jan 18;294(3):1045-1058. doi: 10.1074/jbc.RA118.004471. Epub 2018 Nov 26. PMID: 30478174; PMCID: PMC6341389.
15. Zhang PL, Chen Y, Zhang CH, Wang YX, Fernandez-Funez P. Genetics of Parkinson's disease and related disorders. J Med Genet. 2018 Feb;55(2):73-80. doi: 10.1136/jmedgenet-2017-105047. Epub 2017 Nov 18. PMID: 29151060.

GWAS	Nalls	Blauwendraat	Pankratz	Simon
Year	2019	2019	2012	2009
N_total	482730	34498	8477	5691
N_Case	33674	17996	4238	1713
N_Control	449056	16502	4239	3978
Percent_F_of_Case_and_Control	35.5 and NA	39.3 and NA	35.8 and 55.9	35.9 and 44.1
Age	21 - 84	20 - 89	44 - 77	20 - 87
Reference_Population	UK BioBank EUR	UK BioBank EUR	1000G Phase3	1000G Phase3
N_SNPs_Analyzed	17,510,618	6,813,423	98,200	453,217
N_GenomicRiskLoci	24	1	2	4
N_LeadSNPs	32	2	3	4
N_IndSigSNPs	83	2	3	5
N_CandidateSNPs	7694	119	3923	2920
N_GWASTagged_CandidateSNPs	4703	97	32	23

Supplementary Table 1: This table lists all the GWAS used in analysis for this study. Details about each study are also listed in the table.

GenomicLocus	uniqID	rsID	chr	pos	p	start	end	LeadSNPs	GWAS
1	1:152192927:C:G	rs147288664	1	152192927	6,33E-11	151832302	152277663	rs147288664	Nalls
2	1:155135036:A:G	rs35749011	1	155135036	5,02E-27	154913723	156154860	rs35749011;rs35643925	Nalls
3	1:205656453:C:G	rs823106	1	205656453	4,10E-07	205642390	205757824	rs823106;rs823116	Nalls
4	2:135537119:G:T	rs6741007	2	135537119	2,09E-09	135343934	135583585	rs6741007	Nalls
5	2:169119609:C:G	rs4613239	2	169119609	6,22E-10	169091942	169159689	rs4613239	Nalls
6	3:58218352:A:G	rs4488803	3	58218352	1,08E-05	58218352	58225974	rs4488803	Nalls

7	3:182760073:G:T	rs10513789	3	182760073	3,19E-10	182704808	182860042	rs10513789	Nalls
8	4:951947:C:T	rs34311866	4	951947	7,97E-20	843630	991192	rs34311866	Nalls
9	4:15737348:A:G	rs4698412	4	15737348	7,05E-11	15598051	15744576	rs4698412	Nalls
10	4:77183300:A:C	rs7695720	4	77183300	1,53E-06	77132360	77219402	rs7695720	Nalls
11	4:90666041:C:T	rs356203	4	90666041	3,01E-38	90513519	91307991	rs79956144;rs356203;rs10433953;rs1372518	Nalls
12	5:60345424:G:T	rs75646569	5	60345424	5,62E-10	60011636	60493944	rs4699965;rs75646569	Nalls
13	6:32561334:C:G	rs35265698	6	32561334	3,93E-08	32345007	32613360	rs35265698	Nalls
14	7:23245569:A:G	rs858295	7	23245569	3,83E-06	23110019	23407240	rs858295	Nalls
15	8:16697579:G:T	rs620490	8	16697579	6,46E-07	16695418	16739127	rs620490	Nalls
16	10:121410917:C:T	rs144814361	10	121410917	9,07E-08	121295884	121710488	rs144814361	Nalls
17	11:133764666:C:G	rs329647	11	133764666	1,94E-07	133764405	133819417	rs329647	Nalls
18	12:40734202:A:G	rs34637584	12	40734202	1,33E-09	40385910	41506535	rs1491942;rs34637584;rs75505347	Nalls
19	12:123326598:G:T	rs10847864	12	123326598	9,81E-10	123296294	123326598	rs10847864	Nalls
20	15:61993702:A:G	rs4774417	15	61993702	4,63E-05	61991478	61997599	rs4774417	Nalls
21	16:30923602:A:T	rs12934900	16	30923602	4,33E-08	30801183	31155458	rs12934900	Nalls
22	17:16010920:A:G	rs4566208	17	16010920	3,88E-05	15873275	16225506	rs4566208	Nalls
23	17:44095467:C:T	rs58879558	17	44095467	1,36E-18	43460181	44863413	rs58879558	Nalls
24	18:40672964:A:G	rs4588066	18	40672964	4,45E-06	40672964	40741535	rs4588066	Nalls
1	4:90666041:C:T	rs356203	4	90666041	2,35E-06	90619732	90788196	rs356203;rs983361	Blauwendraat
1	4:90736113:C:T	rs1442144	4	90736113	9,64E-06	90682504	90788196	rs1442144	Pankratz
2	17:44828931:A:G	rs199533	17	44828931	2,58E-08	43463493	44863413	rs413778;rs199533	Pankratz
1	4:90678541:A:G	rs2736990	4	90678541	5,69E-06	90607126	90779710	rs2736990	Simon

2	16:61977449:A:G	rs3784847	16	61977449	1,66E-06	61948150	61990359	rs3784847	Simon
3	17:21717727:G:T	rs4889730	17	21717727	2,83E-08	21702666	21739644	rs4889730	Simon
4	17:44859144:C:T	rs415430	17	44859144	4,50E-05	43859640	44865603	rs415430	Simon

Supplementary Table 2: This table lists all the genomic risk loci identified used in this study. Chromosomal start and end positions are listed for each locus, as well as the lead SNP in the locus and the GWAS the locus was identified with.